Van der Waals surface graphs and molecular shape

David C. Whitley

School of Mathematical Studies and Centre for Molecular Design, Portsmouth University,
Halpern House, Hampshire Terrace, Portsmouth PO1 2QF, UK
E-mail: whitley@sms.port.ac.uk

A *van der Waals surface graph* is the graph defined on a van der Waals surface by the intersections of the atomic van der Waals spheres. A *van der Waals shape graph* has a vertex for each atom with a visible face on the van der Waals surface, and edges between vertices representing atoms with adjacent faces on the van der Waals surface. These are discrete invariants of three-dimensional molecular shape. Some basic properties of van der Waals surface graphs are studied, including their relationship with the Voronoi diagram of the atom centres, and a class of molecular embeddings is identified for which the dual of the van der Waals surface graph coincides with the van der Waals shape graph.

1. Introduction

This paper develops some ideas for describing the three-dimensional shape of molecules which were introduced, informally, in [19]. In contrast with conventional, geometrical descriptions of molecular shape, the approach adopted here combines a topological description of the molecular surface, in the spirit of work by Mezey [13], with the properties of a certain graph defined on the molecular surface. The topological basis of this method provides a flexibility encompassing the intuitive idea that molecules with similar geometrical positions have identical shapes, while the molecular surface graph provides a discrete characterisation of the shape of the molecule. Thus we have a description of molecular shape which is both robust and concise. A major application of these ideas lies in conformational analysis: the study of the various shapes adopted by a single, flexible molecule. The discrete nature of the molecular surface graphs leads to a finite classification of conformational shapes, and enables efficient database searches to be carried out for substructures with a specified shape.

The usual approach to molecular shape is based on a geometrical specification of the way in which a molecule is embedded in three-dimensional space. From a set of admissible embeddings of a molecule, those which may be rigidly superimposed are identified, and the resulting set of geometrically distinct embeddings, the *conformation space*, is the basic object describing the possible shapes of the molecule [5]. However, this geometrical definition of shape is rather restrictive, and the conformation space an unnecessarily large and complex object for many purposes. In particular, molecular

embeddings which differ only slightly, but cannot be rigidly superimposed, represent distinct points in the conformation space, while intuitively they represent the same shape.

A weaker notion of equivalence between molecular embeddings, partitioning the conformation space into sets of conformations with the same shape, and providing an objective definition of the boundaries between different shapes is therefore required; and if small, continuous changes in atomic positions are not to change the shape of a molecule, then topology provides the natural arena in which to work. The easiest way to formulate the problem in topological terms is to associate with each molecular embedding a molecular surface, and consider two embeddings to represent the same shape if the corresponding surfaces are topologically equivalent; that is, if the surfaces can be continuously deformed into one another. However, many molecular surfaces are spheres and have the same topological type, so this purely topological definition of shape is too weak for an adequate description of molecular shape. If molecular surfaces are to be used to discriminate between different molecular shapes, non-topological properties of the surfaces must be incorporated into the theory.

There are, of course, several ways to define a molecular surface, but the simplest is based on the classical description of a molecule as a union of three-dimensional balls representing the atoms. In this case the molecular surface, the van der Waals surface, is a union of spherical polygons formed by the visible parts of the atomic surfaces. The ideas developed below follow from the simple observation that the intersections between the visible faces of the atomic spheres form a graph, the *van der Waals surface graph*, drawn on the molecular surface. A small change in the conformation of a molecule will not, in general, change the van der Waals surface graph, so these graphs are sufficiently stable to accommodate the notion that shape is unaffected by small conformational variations. On the other hand, larger conformational changes alter the arrangement of the visible atomic faces on the van der Waals surface, and hence change the graph. Van der Waals surface graphs therefore provide basic invariants of conformational shape. A formal development of these ideas is presented in sections 2–6.

In the case where the atoms have equal radii, the van der Waals surface graph is closely related to the Voronoi diagram of the atom centres [15]. In section 7, this relationship is used to describe how the van der Waals surface graph changes as the radii vary, and to establish, when the radii are large enough, a connection between the van der Waals surface and the convex hull of the atom centres.

Instead of working directly with the van der Waals surface graphs, it is often convenient to use their geometric duals, which are related to the familiar molecular structure graphs where vertices represent atoms and edges represent bonds. For the simplest molecules, where each atom has a single face on the van der Waals surface, and bonded atoms have adjacent faces, the structure graph is a subgraph of the surface graph dual. The dual graph consists of the structure graph together with extra edges between non-bonded atoms which have van der Waals contacts on the molecular surface. Thus the surface graph dual provides a generalisation of the structure graph, supplementing

the usual connectivity data with information about the three-dimensional shape of the molecule.

For more complex molecules, where atoms may have more than one face on the van der Waals surface, the surface graph dual has more vertices than the structure graph. However, the vertices corresponding to each atom may be identified to obtain a *van der Waals shape graph*, which has a single vertex for each atom, and edges joining pairs of atoms whose van der Waals spheres intersect on the molecular surface (see section 8). This reduction involves some loss of information, and some nice mathematical properties of the surface graphs are lost. (In general, van der Waals surface graphs are trivalent, so their duals are triangulations.) Nevertheless, the attraction of working with graphs which have one vertex per atom is compelling, and empirical studies of a small ring system [19] show that, in this particular case, the van der Waals shape graphs discriminate precisely between the major conformational classes.

These methods provide, quite deliberately, fairly rough descriptions of molecular shape. They suffice, for instance, to separate boats from chairs in cyclohexane (see section 5), but say nothing about variations in shape within these classes. The reader should be aware at the outset that these techniques will only discriminate between conformations with different van der Waals interactions. Situations where part of a molecule is free to rotate about a torsion angle, for example, without making or breaking van der Waals contacts with the rest of the structure, cannot be distinguished this way. The variations in shape detected here are those dictated by short range van der Waals forces rather than longer range electrostatic ones; hence this is a theory of "steric", rather than "electrostatic" shape.

The calculation of a van der Waals surface graph requires a dissection of the van der Waals surface into its constituent spherical faces, and an algorithm to accomplish this is provided in the appendix.

2. Molecular graphs

As the following discussion relies heavily on the language of graph theory, we begin by recalling some basic terminology (see, for example, [9,20]). A *graph G* is a pair (V, E), where V is a non-empty, finite set of *vertices*, and E, the set of *edges* of G, is a finite family of unordered pairs of vertices. A graph G is *connected* if any two vertices are joined by a sequence of edges. An edge joining a vertex to itself is called a *loop*, and if two vertices are joined by more than one edge, these edges are called *multiple edges*. A graph with no loops is a *multigraph*, and a graph with neither loops nor multiple edges is a *simple graph*. A graph G is *embedded in a surface S* if it is drawn on S without crossings; that is, if the edges of G intersect only at their common vertices. In this case a connected component of $S \setminus G$ is called a *face* of G.

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are *isomorphic* if there is a one-to-one correspondence (a *bijection*) $\theta : V_1 \to V_2$ so that the number of edges joining any two vertices in G_1 is the same as the number of edges joining the corresponding

vertices in G_2. An isomorphism of a graph G with itself is called an *automorphism* of G, and the set of all automorphisms of G forms a group, the *automorphism group* of G, denoted Aut(G).

Most of the graphs we encounter are labelled in a natural way, and we use the following notion of a labelled graph which is slightly more general than usual. Given a finite set of labels L, an *L-labelled graph* is a triple $G = (V, E, \lambda)$ consisting of a graph (V, E) together with a labelling $\lambda \colon V \to L$. We emphasise that λ is not assumed to be a bijection. Several vertices of G may have the same label, and some elements of L may not in fact occur as vertex labels. If λ *is* a bijection, then an L-labelled graph is a *labelled graph* in the standard sense. Two L-labelled graphs G_1 and G_2 are isomorphic if there is an isomorphism $\theta \colon G_1 \to G_2$ which preserves the labelling.

Although this is not the main reason for introducing them, L-labelled graphs provide a convenient way to describe molecular graphs. Let $\mathbb{T} = \{H, C, N, \ldots\}$ denote the set of elements in the periodic table. A *molecular graph* G is a $\mathbb{T}$-labelled graph (A, B, c), where (A, B) is a connected graph with vertex set A (the atoms), edge set B (the bonds), and the labelling $c \colon A \to \mathbb{T}$ describes the molecular composition of G. The family $\{c(a) \colon a \in A\}$ is the molecular formula of G. Two molecular graphs G_1 and G_2 are equivalent if they are isomorphic as $\mathbb{T}$-labelled graphs. For present purposes, we consider a *molecule* to be an equivalence class of molecular graphs.

3.   Conformation space

The nature of the molecular graphs themselves is addressed by the classical enumeration problems of chemical graph theory [17]. Here the aim is to describe their shape in three-dimensional Euclidean space $\mathbb{E}^3$, and we formalise the problem along the lines of Crippen and Havel [5].

An *embedding* of a molecule M is a map $\phi \colon G \to \mathbb{E}^3$, where G is a molecular graph representing M. The chemistry of any particular situation determines a set $\Phi(M)$ of *admissible* embeddings of M. Depending on the problem, an admissible embedding may require fixed bond lengths and angles, the mimimisation of a potential energy function, or just the satisfaction of some bound on the potential energy. In general, all we can say is that, since two atoms cannot be in the same place at the same time, these embeddings are injective: if $a \neq a'$ then $\phi(a) \neq \phi(a')$. We will not worry about the definition of admissibility here. All we assume is that $\Phi(M)$ is defined in such a way that it is independent of the choice of the molecular graph G representing M.

The first task when deciding if two embeddings represent the same shape is to identify those which can be rigidly superimposed. Let R^+ and R^- denote the orientation-preserving and orientation-reversing rigid body transformations of $\mathbb{E}^3$, and define an equivalence relation on $\Phi(M)$ by $\phi_1 \sim \phi_2$ if $\phi_1 = \rho \circ \phi_2$ for some $\rho \in R^+$. An equivalence class of embeddings $\chi = [\phi]$ is called a *conformation* of M, and the set of equivalence classes $\mathcal{C}(M) = \Phi(M)/\!\sim$ is the *conformation space* of M. Two conformations χ_1 and χ_2 are *mirror images* if they may be represented by embeddings ϕ_1 and ϕ_2 with $\phi_1 = \rho \circ \phi_2$ for some $\rho \in R^-$.

The conformation space $\mathcal{C}(M)$ is the object underlying many studies of molecular shape but, except in a few cases, its structure is mainly unknown. In general, it is also a rather large space: if all embeddings are admissible, and the molecule has n atoms, then $\mathcal{C}(M)$ has dimension $3n - 6$. This situation is studied in [11], especially for the case $n = 3$. Using a more restrictive definition of admissibility, Crippen and Havel [5] show that for hexagonal rings with fixed bond lengths and angles, $\mathcal{C}(M)$ is the union of two points (two mirror-image chairs) and a circle (a pseudo-rotation of twisted boats).

4. Van der Waals surfaces

As a first step towards a topologically based theory of shape, we associate with each molecular embedding a van der Waals surface, as follows. Assign a radius $r(t)$ to each atom type $t \in \mathbb{T}$. Given an embedding $\phi : G \rightarrow \mathbb{E}^3$ of a molecule M with molecular graph $G = (A, B, c)$, represent each atom $a \in A$ by a closed, three-dimensional ball $B(a)$ centred at $\phi(a)$ with radius $r(c(a))$. The union of these balls

$$\mathcal{M} = \bigcup \big\{ B(a) \colon a \in A \big\}$$

is a solid model of M, and the van der Waals surface $\mathcal{S}$ is its topological boundary

$$\mathcal{S} = \mathrm{bd}(\mathcal{M}) = \mathrm{cl}(\mathcal{M}) \setminus \mathrm{int}(\mathcal{M}),$$

where $\mathrm{cl}(\mathcal{M})$ and $\mathrm{int}(\mathcal{M})$ denote the closure and interior of $\mathcal{M}$, respectively. Thus $\mathcal{S}$ is the set of points of $\mathcal{M}$ which do not lie in the interior of any of the atomic balls $B(a)$. It is a piecewise smooth surface consisting of a union of spherical polygons, each of which is the visible part of an atomic van der Waals sphere $S(a)$, where $S(a)$ is the boundary of $B(a)$.

A van der Waals surface $\mathcal{S}$ is a compact, orientable surface with no boundary, so the topological type of each of its connected components is determined by a single integer invariant, the genus, or number of handles [12]. Although the molecular graph G is assumed to be connected, in exceptional cases there may be cavities within the solid model $\mathcal{M}$, leading to a disconnected $\mathcal{S}$. A procedure for detecting such cavities is described in section A.6 of the appendix (cf. [1]). For a typical molecule, however, $\mathcal{S}$ is connected, in which case a single integer suffices to determine the topology of $\mathcal{S}$. Clearly, this is too crude a description of shape to be of much practical use. A few large ring systems have toroidal van der Waals surfaces, but for many molecules, $\mathcal{S}$ is a surface of genus zero – a topological sphere. Whereas the conformation space is based on a rigid geometrical idea which tends to give too many shapes, this basic topological approach is too flexible and gives too few – in fact, it assigns the same shape to most molecules. To proceed we adopt a path between these two extremes, retaining the underlying flexibility of the topological approach, but augmenting the topology of $\mathcal{S}$ with a certain amount of information determined by the geometry of the embedding.

5.   Van der Waals surface graphs

The additional information we use is provided by the *van der Waals surface graph* $\mathcal{G}$ defined on $\mathcal{S}$ by the intersections of the atomic van der Waals spheres. The faces of $\mathcal{G}$ are the visible parts of the atomic surfaces, the edges are circular arcs forming the intersections between pairs of overlapping atoms, and the vertices are points common to three or more atoms.

To see how these graphs may be useful, consider the van der Waals surfaces for two embeddings of cyclohexane, the chair and twisted boat forms shown in figure 1. Here the hydrogen atoms are omitted, since the underlying shape of these embeddings is determined by the backbone carbon atoms. In both cases, $\mathcal{S}$ is a topological sphere with each atom contributing a single face to $\mathcal{S}$, and $\mathcal{G}$ is the boundary between these faces. To picture $\mathcal{G}$ more easily, use the standard correspondence between spherical and planar graphs: deform $\mathcal{S}$ continuously into a true sphere and stereographically project from a point P in the interior of the face contributed by atom 1 onto the plane tangent to the antipodal point of P. The spherical graph $\mathcal{G}$ projects to a planar graph with one unbounded face corresponding to atom 1 and five bounded faces from the remaining atoms. Figure 2 shows the graphs obtained in this way from the embeddings in figure 1, and we see immediately that the two graphs are isomorphic: they are both the graph of a cube. Hence the surface graphs themselves do not distinguish between the two embeddings.
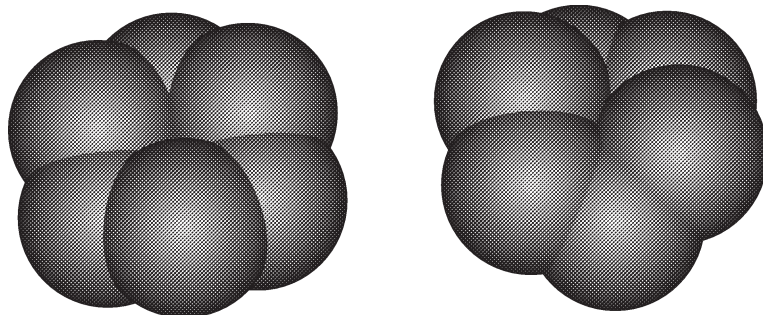

Figure 1. Van der Waals surfaces for chair (left) and twisted boat (right) embeddings of cyclohexane.
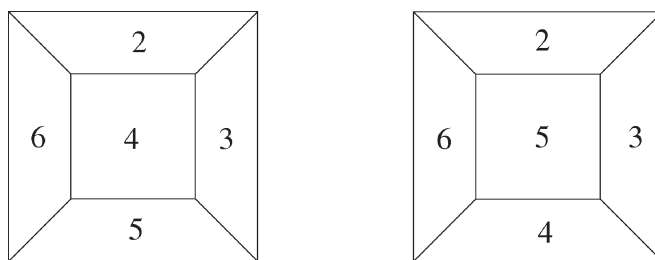

Figure 2. Van der Waals surface graphs for chair (left) and twisted boat (right) embeddings of cyclohexane.

However, if the atoms are numbered, and the faces of the graphs are labelled by the atoms to which they belong, it is clear that the face-labelled graphs in figure 2 are different. (Think of the two surfaces as six-sided dice, and try to match the numbers on corresponding faces.) A crucial point here is that the numbering of the atoms is determined only up to automorphisms of the molecular graph G, in this case, the symmetries of the hexagon. But, as we show at the end of this section, no automorphism of G renumbers the atoms of the chair embedding so that its face-labelled surface graph matches that of the twisted boat embedding. Thus there is a genuine difference between the face-labelled graphs which is not an artefact of the way the atoms are numbered.

It is convenient at this stage to replace the surface graph $\mathcal{G}$ by its geometric dual $\mathcal{G}^*$. The dual graph is constructed by placing a vertex of $\mathcal{G}^*$ in each face of $\mathcal{G}$, and joining two vertices of $\mathcal{G}^*$ if the corresponding faces of $\mathcal{G}$ share a common edge. This process transfers the face labels of $\mathcal{G}$ to vertex labels of $\mathcal{G}^*$. In the cyclohexane example, each atom provides a unique face of $\mathcal{S}$, so there is a bijection between the vertices of $\mathcal{G}^*$ and the vertices of the molecular graph $G = (A, B, c)$. Thus $\mathcal{G}^*$ is a labelled graph (in the usual sense) with label set A. In other cases, however, an atom may contribute more than one face to $\mathcal{S}$, or be buried inside the molecule with no visible faces. Therefore several vertices of $\mathcal{G}^*$ may have the same label, and some members of A may not occur as labels of $\mathcal{G}^*$, so $\mathcal{G}^*$ is an A-labelled graph.

Thus an embedding ϕ of a molecular graph $G = (A, B, c)$ determines a surface $\mathcal{S}$ and an A-labelled graph $\mathcal{G} = (V, E, \lambda)$ embedded in $\mathcal{S}$. Suppose that ϕ_1 and ϕ_2 are two such embeddings with corresponding surfaces $\mathcal{S}_1, \mathcal{S}_2$, graphs $\mathcal{G}_1^*, \mathcal{G}_2^*$, and labellings λ_1, λ_2. The previous example motivates a three-stage procedure for deciding if ϕ_1 and ϕ_2 represent the same shape. First, the embeddings should have topologically equivalent van der Waals surfaces; that is, there should be a homeomorphism (a continuous map with a continuous inverse) between $\mathcal{S}_1$ and $\mathcal{S}_2$. Second, the surface graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ should be isomorphic. If both these conditions hold, the isomorphism between $\mathcal{G}_1$ and $\mathcal{G}_2$ can be extended to a homeomorphism between $\mathcal{S}_1$ and $\mathcal{S}_2$, so we may assume that there is a homeomorphism $\theta : \mathcal{S}_1 \to \mathcal{S}_2$ whose restriction to $\mathcal{G}_1^*$, which we also denote by θ, is an isomorphism $\theta : \mathcal{G}_1^* \to \mathcal{G}_2^*$. Finally, the map θ should preserve the labellings λ_1 and λ_2, modulo automorphisms of the molecular graph G. This provides our basic definition.

Definition 1. Two embeddings $\phi_1, \phi_2 \in \Phi(M)$ have the same *van der Waals shape* if there is an orientation-preserving homeomorphism $\theta : \mathcal{S}_1 \to \mathcal{S}_2$ so that $\theta : \mathcal{G}_1^* \to \mathcal{G}_2^*$ is an isomorphism, and an automorphism $\varphi : G \to G$ so that $\lambda_2(\theta(v)) = \varphi(\lambda_1(v))$ for each vertex v of $\mathcal{G}_1^*$.

This determines an equivalence relation $\approx$ on $\Phi(M)$ in the obvious way: $\phi_1 \approx \phi_2$ if ϕ_1 and ϕ_2 have the same van der Waals shape. A *van der Waals shape* is an equivalence class of embeddings $[\phi]$. Given an embedding $\phi \in \Phi(M)$, any reflection r in a plane in $\mathbb{E}^3$ determines a mirror image embedding $\phi' = r \circ \phi$. An embedding ϕ
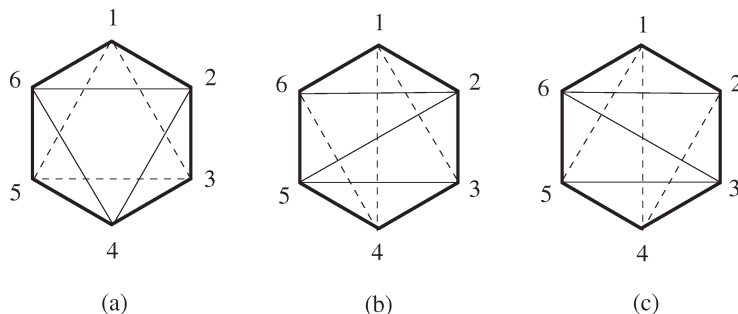
Figure 3. Van der Waals shapes for chair (a) and twisted boat (b), (c) embeddings of cyclohexane.

has a *symmetric* van der Waals shape if $[\phi] = [\phi']$, and an *enantiomeric* van der Waals shape if $[\phi] \neq [\phi']$. Clearly, this is independent of the choice of reflection r. If ϕ_1 and ϕ_2 represent the same conformation then they have the same van der Waals shape, with the equivalence θ given by the rigid body transformation relating ϕ_1 and ϕ_2. Hence $\approx$ may be considered as an equivalence relation on the conformation space, partitioning $\mathcal{C}(M)$ into sets of conformations with the same van der Waals shape.

The van der Waals shapes for the chair and twisted boat embeddings of cyclohexane are shown in figure 3. The dual graphs $\mathcal{G}^*$ are drawn with their vertices placed around the equator of $\mathcal{S}$, and edges in the upper and lower hemispheres of $\mathcal{S}$ are drawn with solid and dashed lines, respectively. Any automorphism of G preserves the 6-cycle shown in bold around the equator of $\mathcal{S}$, so any isomorphism θ as in definition 1 preserves the subgraphs of $\mathcal{G}^*$ in each hemisphere. Hence the graphs in figure 3 represent different van der Waals shapes. The chair embedding has a symmetric van der Waals shape (figure 3(a)), and the twisted boat embedding has one of a pair of enantiomeric van der Waals shapes (figure 3 (b) and (c)).

The construction of the van der Waals surface and the resulting van der Waals shape depends on the choice of the atomic radii. Earlier work [19] showed that while the van der Waals surface graphs of a collection of embeddings of a small ring molecule changed with the choice of radii, the sets of embeddings having the same van der Waals shape did not. Hence the atomic radii were not crucial in using the van der Waals shapes to classify the embeddings. Studies of larger molecules indicate that their van der Waals shapes may be more sensitive to the choice of atomic radii, but it is intuitively clear that the van der Waals surface graph is constant over small ranges of radii, as the following section makes precise.

6. Transversal embeddings

The van der Waals surface graph $\mathcal{G}$ is determined by the atomic coordinates and radii, and an algorithm to carry out this calculation is described in the appendix. A brief outline of the algorithm is given below, but the main purpose of this section is to discuss the ways in which the atomic spheres intersect and establish a useful generic

property of the surface graphs. The algorithm for $\mathcal{G}$ is based on a systematic analysis of the intersections between the atomic spheres, but as we are seeking a description of those molecular shapes which are unaltered by small conformational changes, there is no need to consider all possible types of intersection between collections of atomic spheres. It suffices to consider only those intersections which are typical and stable; qualities which are captured formally by the notion of *transversal* intersection [3,8].

For example, two atomic spheres in $\mathbb{E}^3$ may be disjoint, touch tangentially at a single point, or intersect in a circle. The central possibility, tangency, is a non-transversal intersection, while the other intersections are transversal. No small change in the atomic coordinates will qualitatively alter a transversal intersection. But a small perturbation which changes the distances between the atom centres will break a tangency, either pulling the spheres apart, or pushing the centres closer, forcing the spheres to intersect in a circle. Thus transversal intersections between two spheres are stable, and non-transversal ones unstable. Moreover, transversal intersection is the typical case here, since two spheres chosen "at random" would not be expected to produce a tangency.

Extending this example, suppose the two spheres S_i and S_j intersect transversally in a circle C_{ij}, and consider a third sphere S_k. The possible types of intersection between C_{ij} and S_k are empty intersection, tangency, and non-empty intersection where S_k meets C_{ij} in a pair of isolated points. Once again, the special case of tangency is a non-transversal intersection, while the other intersections are transversal. Furthermore, in the latter case, the two points in $C_{ij} \cap S_k$ form the triple intersection $S_i \cap S_j \cap S_k$, and a transversal intersection between either of these points and a fourth sphere is necessarily empty. Thus if the intersections are transversal, two spheres are either disjoint or meet in a circle, three spheres either have no points in common or meet in two isolated points, and four or more spheres have no common points.

Intuitively, the situation described above is the generic case. To make this more precise, define a metric on $\Phi(M)$ by

$$d(\phi_1, \phi_2) = \sum_{a \in A} \left\| \phi_1(a) - \phi_2(a) \right\|,$$

where $\| \cdot \|$ is the Euclidean norm on $\mathbb{E}^3$. Let $\Phi_T(M)$ be the set of admissible embeddings of M for which S_{i_0} intersects $S_{i_1} \cap \cdots \cap S_{i_k}$ transversally, for each k-tuple $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ and each $i_0 \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$, where $S_1, \ldots, S_n$ are the atomic spheres defined by the embedding. An embedding $\phi \in \Phi_T(M)$ is called a *transversal* embedding. It follows from the transversality theorem [8] that $\Phi_T(M)$ is an open and dense subset of $\Phi(M)$, in the topology induced by the metric d. This is the formal statement that transversal embeddings are stable and typical. (For those unfamiliar with the terms, $\Phi_T(M)$ is *open* in $\Phi(M)$ if for each $\phi \in \Phi_T(M)$, any embedding ϕ' with $d(\phi, \phi')$ sufficiently small also lies in $\Phi_T(M)$, while $\Phi_T(M)$ is *dense* in $\Phi(M)$ if given any $\phi \in \Phi(M)$, there is a $\phi' \in \Phi_T(M)$ with $d(\phi, \phi')$ arbitrarily small.)

The algorithm for $\mathcal{G}$ proceeds by considering the circle of intersection C between each pair of overlapping atomic spheres. If no further atoms intersect C, a vertex v is placed arbitrarily on C, and C becomes a loop of $\mathcal{G}$ joining v to itself. If C does intersect other atomic spheres, the points of triple intersection separate C into a number of arcs. The visible arcs, those which do not lie in the interior of an atomic ball, are edges of $\mathcal{G}$; the points bounding visible arcs are vertices of $\mathcal{G}$; and the faces of $\mathcal{G}$ are found by grouping the edges of $\mathcal{G}$ into cycles. The geometrical details are given in the appendix. If at any stage a non-transversal intersection between atomic spheres is discovered, the embedding is perturbed slightly and the algorithm restarted. Notice that each vertex which is a point of triple intersection between atomic spheres lies on three circles of pairwise intersection, and hence three edges meet at such a vertex. These observations are summarised as follows.

Theorem 2. There is an open and dense subset $\Phi_T(M) \subset \Phi(M)$, so that for each embedding $\phi \in \Phi_T(M)$, the connected components of $\mathcal{G}$ are either loops or trivalent subgraphs of $\mathcal{G}$.

Loops in $\mathcal{G}$ fall into two types, depending on whether they form the intersection between a pair of bonded or non-bonded atoms. In the former case, it is already clear that there is van der Waals contact between the two atoms, so $\mathcal{G}$ provides no new information. Indeed, such a loop contributes an edge to $\mathcal{G}^*$ between vertices labelled by the intersecting atoms, which corresponds directly to an edge of G. On the other hand, a loop of $\mathcal{G}$ between non-bonded atoms provides information about the shape of the embedding which is not present in the molecular graph; such a loop contributes an edge to $\mathcal{G}^*$ which has no counterpart in G.

Any loop in $\mathcal{G}$ may be removed by cutting $\mathcal{S}$ along the loop to produce a surface with two holes, and then glueing discs into the holes to regain a surface without boundary. This will change the topology of $\mathcal{S}$, and may separate $\mathcal{S}$ into a number of surfaces. For instance, if $\mathcal{S}$ is a sphere this operation will divide $\mathcal{S}$ into two spheres, while removing a loop from a torus in this way will change $\mathcal{S}$ into a sphere. If, after removing the loops, $\mathcal{G}$ is still disconnected, we may separate the components of $\mathcal{G}$ by simple closed curves on $\mathcal{S}$ and remove these in the same manner. In this way, $\mathcal{S}$ can be decomposed into a number of surfaces, on each of which $\mathcal{G}$ is a connected, trivalent multigraph with labelled faces; or, equivalently, the dual graph $\mathcal{G}^*$ is an A-labelled triangulation.

7.   Voronoi diagrams

The previous section addressed the structure of $\mathcal{G}$ for a typical embedding, given a fixed set of atomic radii. As the van der Waals radii are determined empirically and apt to vary from one application to another, it is useful to know how $\mathcal{G}$ changes as the

radii vary for a fixed embedding. In this section, we examine the case where the radii are identical for each atom, by studying the *Voronoi diagram*

$$\mathcal{V}(\phi) = \big\{V(a)\big\}_{a \in A}$$

where

$$V(a) = \big\{p \in \mathbb{E}^3 \colon \big\|p - \phi(a)\big\| \leqslant \big\|p - \phi(b)\big\| \ \forall b \in A \setminus \{a\}\big\}$$

is the *Voronoi region* of atom a. For the general theory of Voronoi diagrams, see [2,6,15]. The Voronoi diagram is a polyhedral partition of $\mathbb{E}^3$ into convex regions whose interior points lie closer to one atom centre than to any other. The boundary of a Voronoi region consists of two-dimensional *Voronoi faces*; the boundary of a Voronoi face consists of *Voronoi edges*; and boundary of a Voronoi edge consists of *Voronoi vertices*. Voronoi vertices are equidistant from four or more atom centres; points in a Voronoi edge are equidistant from three or more atom centres; and points in a Voronoi face are equidistant from two or more atom centres.

To see the connection between Voronoi diagrams and van der Waals surface graphs, imagine the atoms growing as spherical wavefronts from their centres at a constant rate. While an atom is sufficiently small, it is disjoint from the others and lies within its own Voronoi region. As the atoms grow larger, they begin to collide and the collision points mark the boundaries of the Voronoi regions. Points where two or more atoms meet lie in Voronoi faces; points where three or more meet lie in Voronoi edges; and points where four or more meet are Voronoi vertices. To simplify things a little, throughout this section we assume the following conditions.

Definition 3. An embedding is *non-degenerate* if

 (i) no four atom centres are co-planar; and

(ii) no five atom centres are co-spherical.

For the moment, it is sufficient to replace (i) by the weaker condition:

(i′) no four atom centres are co-circular.

This implies that points in the interior of a Voronoi edge are equidistant from exactly three atom centres, and (ii) guarantees that Voronoi vertices are equidistant from exactly four atom centres. Condition (i) will be required later.

If the radius of each atom is r, the van der Waals surface $\mathcal{S}_r$ is the level set $D^{-1}(r)$ of the "distance" function $D \colon \mathbb{E}^3 \to \mathbb{R}$ defined by

$$D(p) = \min_{a \in A} \big\|p - \phi(a)\big\|.$$

That is,

$$\mathcal{S}_r = \big\{p \in \mathbb{E}^3 \colon \ D(p) = r\big\},$$

and the van der Waals surface graph is the subset of $\mathcal{S}_r$ where the atoms intersect:

$$\mathcal{G}_r = \mathcal{S}_r \cap \bigcup_{a \in A} \mathrm{bd}\big(V(a)\big).$$

To understand how $\mathcal{G}_r$ changes with r, consider the level sets $D^{-1}(r)$ near the boundaries of the Voronoi regions. For a differentiable function f, changes in the topology of the level sets $f^{-1}(r)$ occur only at critical points of f (where $f'(p) = 0$), and are described by Morse theory [14]. Here D is not differentiable at the atom centres or on the boundaries of the Voronoi regions, but the qualitative behaviour of D is similar to that of a differentiable function. The following observations are straightforward extensions of results obtained by Siersma [16] for two-dimensional Voronoi diagrams.

The atom centres are topological minima of D. In the interior of a Voronoi region $V(a)$, we have $D(p) = \|p - \phi(a)\|$, so any $p \neq a$ is a regular point of D ($D'(p) \neq 0$), and near p, the level sets $D^{-1}(r)$ are subsets of spheres of radius r centred at $\phi(a)$. Now consider the possible positions of a point $p \in \mathrm{bd}(V(a))$.

(i) *p lies in the interior of a Voronoi face F*. The face F is the common boundary between two Voronoi regions $V(a)$ and $V(b)$, and is orthogonal to the line segment $L = \overline{\phi(a)\phi(b)}$. If $p = F \cap L$ then p is a topological saddle point of D of index 1, and near p, the level sets of D are homeomorphic to those of the function $(x, y, z) \mapsto x^2 + y^2 - z^2$. Such a point marks the stage where two spheres first intersect as the radii grow, with $\mathcal{G}_r$ changing from the empty set to a loop (see figure 4(a)). If $p \neq F \cap L$ then near p there is a homeomorphic change of coordinates converting D to a linear function, so p is a topological regular point of D. In this case, there is no change in $\mathcal{G}_r$ as r varies: $\mathcal{G}_r = D^{-1}(r) \cap F$ remains a circular arc forming part of the intersection between spheres of radius r centred at $\phi(a)$ and $\phi(b)$.

(ii) *p lies in the interior of a Voronoi edge E*. Points in E are equidistant from three atom centres. The circumcentre of the atom centres is a topological saddle point of index 2, and near p, the level sets of D are homeomorphic to those of $(x, y, z) \mapsto x^2 - y^2 - z^2$ (see figure 4(b)). This situation occurs when three spheres form a point of triple intersection. The graph $\mathcal{G}_r$ changes from three disjoint circular
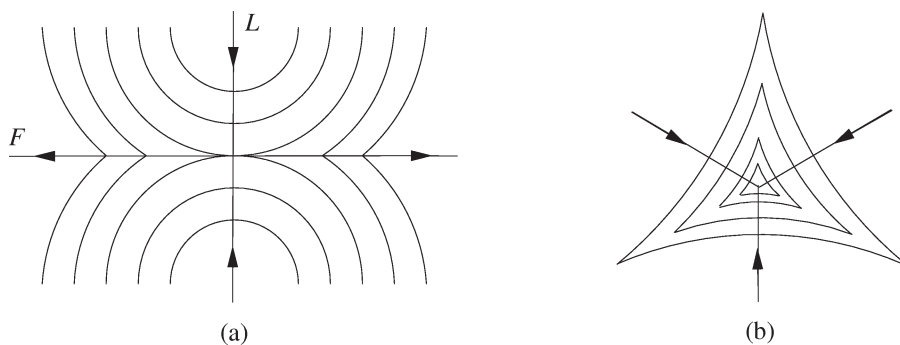

Figure 4. Level sets of D in the plane defined by the atom centres near topological saddles of index 1 (a) and 2 (b). D increases in the direction of the arrows and perpendicular to the plane of the atom centres.

arcs to six arcs joined at two triple points. If p is not the circumcentre of these atom centres then $\mathcal{G}_r$ has a triple point at p, and there is no local change in $\mathcal{G}_r$ as r varies.

(iii) *p is a Voronoi vertex.* The vertex p is equidistant from four atom centres defining a tetrahedron T. If $p \in \text{int}(T)$ then p is a topological maximum of D, and near p, the level sets of D are homeomorphic to those of $(x, y, z) \mapsto -x^2 - y^2 - z^2$. This situation occurs when the solid model $\mathcal{M}$ contains a spherical cavity bounded by four spheres, and the cavity is filled as the atoms grow. The graph $\mathcal{G}_r$ on the surface of the cavity changes from a tetrahedron to the empty set as r increases.

If $p \notin \text{int}(T)$ there are two possibilities, depending on the position of p with respect to the planes containing the faces of T. Consider T and its circumscribing sphere Σ. If T is a regular tetrahedron then $p \in \text{int}(T)$ is a maximum of D as described above. Now suppose the vertices of T are allowed to move on Σ, while Σ remains fixed. There are two ways in which p may leave T as the vertices move on Σ: through a face or through an edge. The first case is shown in figure 5(a). Here D increases as we approach p along three of the Voronoi edges through p, and decreases along the fourth. The vertex p is a topologically regular point of D, and a triangular face of $\mathcal{G}_r$ collapses to a triple point as the level sets of D pass through p (see figure 6(a)). The second case is shown in figure 5(b). Here D increases as we approach p along two of the Voronoi edges through p, and decreases along the other two. Again, p is a topologically regular point of D, and two triple points of $\mathcal{G}_r$ meet in a quadruple point as the level sets of D pass through p, causing the rearrangement of the faces of $\mathcal{G}_r$ shown in figure 6(b).

Thus $\mathcal{G}_r$ changes at Voronoi vertices of ϕ and topological saddles of D. As there are only finitely many such points, the atomic radius can be chosen so that $\mathcal{M}_r$ contains them all, and then $\mathcal{G}_r$ is constant for all larger radii. In fact, it is sufficient that $r > r_0$, where r_0 is chosen so that for each $a \in A$ all Voronoi vertices in $V(a)$ lie in $B_{r_0}(a)$. This follows by considering the Delauney tesselation [15], the dual
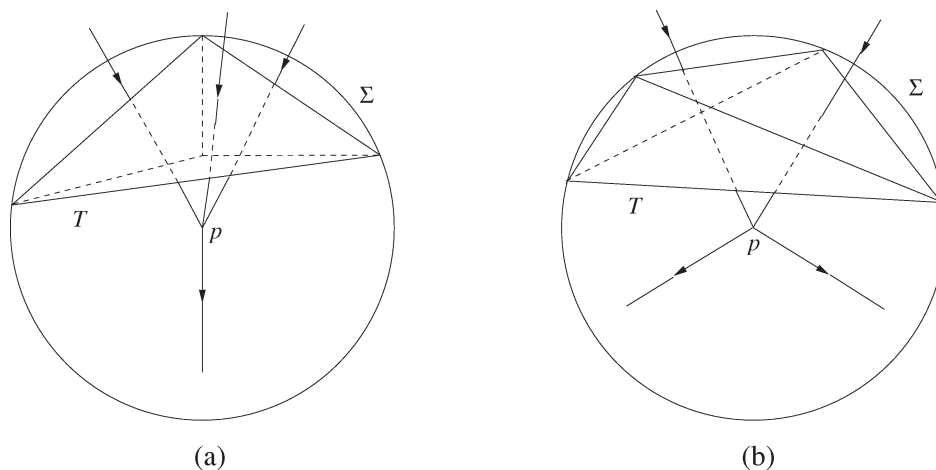

Figure 5. Possible arrangements of Voronoi edges near topologically regular Voronoi vertices. The Voronoi edges are contained in lines orthogonal to the faces of T passing through their circumcentres.
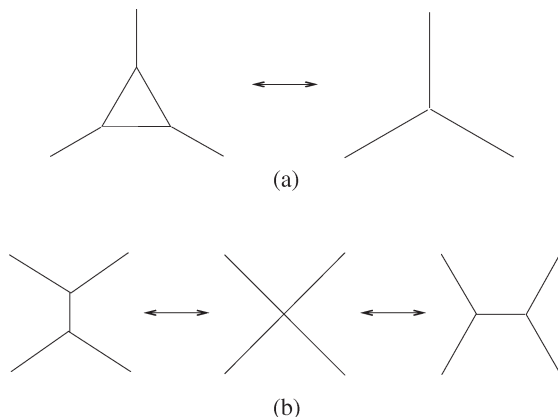
Figure 6. Changes in $\mathcal{G}_r$ near topologically regular Voronoi vertices.

of the Voronoi diagram. Let $q_1, \ldots, q_m$ be the Voronoi vertices of a non-degenerate embedding ϕ, and let $V(a_{i_1}), \ldots, V(a_{i_4})$ be the four Voronoi regions containing q_i. The *Delauney tesselation* $\mathcal{D}(\phi)$ is defined by

$$\mathcal{D}(\phi) = \{T_i\}_{i=1}^m,$$

where T_i is the convex hull of $\{\phi(a_{i_1}), \ldots, \phi(a_{i_4})\}$. As ϕ is non-degenerate, each T_i is a tetrahedron with vertices $\phi(a_{i_1}), \ldots, \phi(a_{i_4})$. It follows from the definition [15] that

$$\bigcup_{i=1}^m T_i = \mathrm{ch}(\phi),$$

where $\mathrm{ch}(\phi)$ is the convex hull of the atom centres. The faces, edges and vertices of the Delauney tetrahedra contained in $\mathrm{bd}(\mathrm{ch}(\phi))$ are called *external* Delauney faces, edges and vertices. The boundary of $\mathrm{ch}(\phi)$ is a topological sphere, triangulated by the graph $\mathcal{G}_{\mathrm{ch}}$ defined by the external Delauney edges. The solid model $\mathcal{M}_r$ is homeomorphic to a three-dimensional ball, being the union of $\mathrm{ch}(\phi)$ with the atomic spheres centred at external Delauney vertices. In particular, $\mathcal{M}$ contains no cavities.

By the choice of r_0 we have $V(a) \subset B_r(a)$ for each bounded Voronoi region, so each face, edge and vertex of $\mathcal{G}_r$ is the intersection of $\mathcal{S}_r$ with an unbounded Voronoi region, face and edge, respectively. But the latter are in one-to-one correspondence with the external Delauney vertices, edges and faces, establishing the following result.

Theorem 4. Let ϕ be a non-degenerate embedding of a molecule M. Choose r_0 so that for each $a \in A$ all the Voronoi vertices in $V(a)$ are contained in $B_{r_0}(a)$. Then, for each $r > r_0$,

(i) $\mathcal{S}_r$ is homeomorphic to $\mathrm{bd}(\mathrm{ch}(\phi))$; and

(ii) $\mathcal{G}_r^*$ is isomorphic to $\mathcal{G}_{\mathrm{ch}}$.

If the atomic radii differ, the function D may be replaced by

$$D(p) = \min_{a \in A} \omega_a \|p - \phi(a)\|$$

with the weights $\omega_a = R/r(a)$ chosen so that $\mathcal{S} = D^{-1}(R)$ once again. This corresponds to the atoms growing at different rates as R increases, and the Voronoi regions are replaced by spherical regions forming the Apollonius model [2]. In this case, the atoms with the largest radius $r(a)$ contain the smaller atoms when R is large.

8. Van der Waals shape graphs

In this section, we return to the general case where the atomic radii vary from atom to atom. Since each atom may have several faces on $\mathcal{S}$, there is no general bound on the number of vertices of $\mathcal{G}^*$. But it is often convenient to work with graphs having (at most) one vertex per atom, and to this end, we introduce the notion of a *van der Waals shape graph*, denoted $\mathcal{G}_{\mathcal{S}}$, which is obtained from $\mathcal{G}^*$ by collapsing all the vertices of $\mathcal{G}^*$ coming from the same atom into a single vertex of $\mathcal{G}_{\mathcal{S}}$.

Definition 5. The *van der Waals shape graph $\mathcal{G}_{\mathcal{S}}$* of an embedding $\phi \in \Phi(M)$ has one vertex for each atom with a visible face on $\mathcal{S}$, and two vertices of $\mathcal{G}_{\mathcal{S}}$ are joined by an edge if the corresponding atoms have a common edge on $\mathcal{S}$.

Note that $\mathcal{G}_{\mathcal{S}}$ is a simple graph. To investigate the relationship between $\mathcal{G}_{\mathcal{S}}$ and $\mathcal{G}^*$ we first isolate some straightforward properties of molecular embeddings. Recall that a subgraph H of a graph K is a *spanning* subgraph if H and K have the same vertex set, and in this case, we say that H *spans* K.

Definition 6. Let ϕ be an embedding of a molecular graph G with van der Waals surface $\mathcal{S}$ and van der Waals surface graph $\mathcal{G}$.

 (i) ϕ is *triangular* if $\mathcal{G}^*$ is a triangulation of $\mathcal{S}$;

 (ii) ϕ is *spherical* if $\mathcal{S}$ is homeomorphic to a sphere;

(iii) ϕ is *simple* if $\mathcal{G}^*$ is a simple graph;

(iv) ϕ is *basic* if ϕ is transversal, triangular, and spherical, and G spans $\mathcal{G}^*$.

The following result lists some immediate consequences of these definitions.

Proposition 7. Let ϕ, G, $\mathcal{S}$ and $\mathcal{G}$ be as in definition 6.

 (i) If ϕ is transversal and $\mathcal{G}$ is connected, then ϕ is triangular.

 (ii) If ϕ is simple and basic, then $\mathcal{G}^*$ is isomorphic to $\mathcal{G}_{\mathcal{S}}$.

(iii) If each pair of atoms has at most one common edge on $\mathcal{S}$, then ϕ is simple.

(iv) If each atom has at least one face on $\mathcal{S}$, and each pair of bonded atoms has at least one common edge on $\mathcal{S}$, then G spans $\mathcal{G}_{\mathcal{S}}$.

(v) If each atom has a unique face on $\mathcal{S}$, and each pair of bonded atoms has a unique common edge on $\mathcal{S}$, then G spans both $\mathcal{G}^*$ and $\mathcal{G}_{\mathcal{S}}$, and $\mathcal{G}^*$ is isomorphic to $\mathcal{G}_{\mathcal{S}}$.

The conditions of proposition 7(iv) are satisfied by embeddings of many small molecules. In this case, we may think of $\mathcal{G}_{\mathcal{S}}$ as being obtained from G by adding edges to G between non-bonded atoms which have contacts on the van der Waals surface. Thus $\mathcal{G}_{\mathcal{S}}$ is a natural extension of G containing information on both the connectivity of the molecule and the three-dimensional shape of the embedding.

Finally, the definition of van der Waals shape is slightly easier for basic embeddings.

Proposition 8. Two basic embeddings ϕ_1 and ϕ_2 with van der Waals surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$ and van der Waals surface graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same van der Waals shape if there is an orientation-preserving homeomorphism $\theta : \mathcal{S}_1 \to \mathcal{S}_2$ so that $\theta : \mathcal{G}_1^* \to \mathcal{G}_2^*$ is an isomorphism (of spherical triangulations) and $\theta \in \mathrm{Aut}(G)$.

Proof. Since basic embeddings are spherical, the condition that the van der Waals surfaces are homeomorphic is automatically satisfied. The van der Waals surface graphs are connected because the embeddings are G-spanning, and together with the transversality condition this ensures that the surface graph duals are triangulations. The G-spanning condition also implies that θ restricted to G plays the role of ϕ in definition 1.   □

9.   Concluding remarks

Van der Waals surface graphs are natural invariants of three-dimensional molecular shape. For simple, basic embeddings their duals coincide with van der Waals shape graphs, which are themselves natural extensions of molecular graphs. Both van der Waals surface graphs and van der Waals shape graphs are straightforward to calculate, and the appendix describes an algorithm that takes $\mathrm{O}(n^3)$ time for an n-atom molecule. The interpretation of the graphs is transparent, and they embody three-dimensional information in a discrete format with all the advantages in terms of storage requirements and processing speed that implies.

There are many potential applications of these graphs. Perhaps the most obvious is that they provide a finite classification of conformational shapes, and a classification of the shapes of small rings based on these ideas will appear elsewhere [18]. It is also clear that several other methods currently used to analyse molecular graphs may be applied to van der Waals shape and surface graphs to obtain results incorporating the three-dimensional shape of a molecular embedding. Leading candidates are connectivity indices [17] in quantitative structure–activity relationships and fingerprints [10] in studies of molecular similarity.

Appendix. Dissecting a van der Waals surface

This appendix describes an algorithm for dissecting a van der Waals surface $\mathcal{S}$ into its spherical faces. This includes calculations of the van der Waals surface graph $\mathcal{G}$, the shape sequences used in [19], and the topological type of the surface. It also detects cavities within the solid model of the molecule (cf. [1]). An outline of this algorithm appeared previously in [19].

The algorithm begins by determining which atoms overlap, then calculates geometric data about the circles of intersection between overlapping atoms which are needed later (section A.2). The vertices of the graph are found from triple intersections of atoms; each triple contributes two vertices (section A.3). The relative locations of the vertices on the circles determines the edges of the graph (section A.4), and the edges are grouped into cycles forming the boundaries of the faces (section A.5). Finally, the faces are assembled in the order determined by the partition and the shape sequence is calculated (section A.6). The calculations require several results from coordinate geometry, and these are collected together in section A.1.

A.1. Geometry

Suppose the molecule consists of atoms $B_1, \ldots, B_n$, where each B_i is a closed ball in $\mathbb{E}^3$ with centre c_i and radius r_i. Let S_i be the the boundary of B_i – the sphere with centre c_i and radius r_i. The statements below follow from elementary geometry.

(A1) $B_i \cap B_j \neq \emptyset \Leftrightarrow \|c_i - c_j\| \leqslant r_i + r_j$.

(A2) $B_i \subset B_j \Leftrightarrow \|c_i - c_j\| \leqslant r_i - r_j$.

(A3) If $|r_i - r_j| < \|c_i - c_j\| \leqslant r_i + r_j$ then $S_i \cap S_j$ is a circle C_{ij} with centre c and radius r, where

$$c = c_j + \lambda(c_i - c_j),$$
$$r = \sqrt{r_j^2 - \lambda^2 \|c_i - c_j\|^2}$$

and

$$\lambda = \frac{\|c_i - c_j\| - r_i^2 + r_j^2}{2\|c_i - c_j\|}.$$

(A4) A basis $\{b_1, b_2\}$ for the plane containing C_{ij} may be chosen as follows. Suppose $c_j - c_i = (x, y, z)$. If $x^2 + y^2 \approx 0$, let $b_1 = (\text{sign}(z), 0, 0)$ and $b_2 = (0, 1, 0)$. Otherwise, let $b_1 = (y, -x, 0)$ and $b_2 = (-xz, -yz, x^2 + y^2)$. This basis is chosen so that $\det(b_1, b_2, c_i - c_j) > 0$; that is, $\{b_1, b_2, c_i - c_j\}$ is a right-handed coordinate system.

(A5) If the basis vectors are normalised to have length r, i.e., $b_i \mapsto rb_i/\|b_i\|$, then the circle C_{ij} is parameterised by $\nu = c + b_1 \sin \theta + b_2 \cos \theta$ $(0 \leqslant \theta \leqslant 2\pi)$. By the choice of the basis in (A5), the circle C_{ij} is oriented counter-clockwise when viewed along the line from c_i to c_j.

(A6) Now consider a third atom B_k. We have

$$C_{ij} \cap \text{int}(B_k) \neq \emptyset \quad \Leftrightarrow \quad \alpha^2 + \beta^2 > \frac{1}{4}\left(r_k^2 - \|c - c_k\|^2 - r^2\right),$$

where $\alpha = b_1(c - c_k)$ and $\beta = b_2(c - c_k)$.

(A7) If $C_{ij} \cap \text{int}(B_k) \neq \emptyset$, then C_{ij} meets S_k in two points, with angles in C_{ij} (in terms of the parameterisation of (A5))

$$\theta_1 = \phi + a\cos\omega \quad (\text{mod } 2\pi),$$
$$\theta_2 = \phi + 2\pi - a\cos\omega \quad (\text{mod } 2\pi),$$

where $\omega = (r_k^2 - \|c - c_k\|^2 - r^2)/2\sqrt{\alpha^2 + \beta^2}$ and $\tan\phi = \alpha/\beta$.

A.2. Circles

First, we calculate a Boolean matrix *overlap* showing which atoms intersect pairwise. This information is used to check that the atoms form a connected set. The atoms are also checked to see if any are completely hidden inside others, and any hidden atoms are removed from further consideration by setting all *overlap* entries relating to such atoms to *false*. The remaining pairwise intersections are counted to give the number of circles of intersection. For each pair of atoms B_i, B_j, $i < j$, with $overlap(i,j) = true$, we calculate and store the centre and radius of the circle C_{ij} from (A3), and the normalised basis vectors from (A4)/(A5).

A.3. Vertices

For each triple $i < j < k$ with

$$overlap(i,j) = overlap(i,k) = overlap(j,k) = true,$$

(A6) is used to decide if $S_i \cap S_j \cap S_k \neq \emptyset$. If this is a triple intersection, the angles of the vertices in each of the circles C_{ij}, C_{ik}, and C_{jk} are calculated from (A7). The six angles must now be divided into two triples, so that all the members of a triple refer to the same vertex. This is done by finding and comparing the coordinates of the vertex given by each angle in (A5). To avoid duplicating any calculations the vertex coordinates are found at the same time as the angles. Finally, the vertices are checked to see if they are hidden by any atoms other than B_i, B_j, and B_k.

A.4. Edges

To find the edges of the graph on the molecular surface, we consider, in turn, each circle of intersection between overlapping atoms.

If there are no vertices on the circle, the circle is a loop forming the intersection between two atoms, and no other atoms meet the circle. Before adding the loop to the list of edges, we check that the loop is visible. Now the whole circle is either visible

or hidden by other atoms; a partially hidden circle would contain vertices and so not be a loop. Thus it is sufficient to check the visibility of a single point on the loop, and we use the point given by (A5) with $\theta = 0$. If this point is visible it is treated as a new vertex, and an edge is added joining the vertex to itself. If the point is hidden the circle is simply ignored.

If there are vertices on the circle, these are ordered by their angles in (A7), and each ordered pair is examined to see if it determines an edge on the molecular surface. We check first that the vertices do not coincide – coincident vertices correspond to non-generic intersections of four or more atoms in a point, and in this case, the molecule is perturbed and the calculation restarted. In the general case, the mid-point of the arc between the two vertices on the circle is calculated using (A5), and if this does not lie inside any other atom, then we have a visible edge joining the two vertices.

The calculation of the van der Waals surface graph, as a collection of vertices and edges, is complete at this stage.

A.5. Faces

To find the faces of $\mathcal{G}$, the edges on each atom are collected into cycles (closed loops formed by sequences of adjoining edges), and the cycles are partitioned into groups bounding a single face using the procedure employed by Connolly [4].

Recall from (A5) that the circles C_{ij} are oriented so that the visible part of S_i lies to the left of C_{ij}, and the visible part of S_j to the right. For a sphere S_k, if the orientation of edges on circles C_{ik} with $i < k$ are retained and the orientations of edges on circles C_{kj} with $k < j$ are reversed, the visible part of S_k will lie to the immediate left of any edge. By construction, the orientations on the edges in a cycle are consistent, and induce an orientation on the cycle. Thus the interior of a cycle may be defined to be the region of the atomic surface to the left of the cycle when traversing the cycle in the direction of the orientation.

A necessary condition for two cycles A and B to belong to the same face is that they lie in each others interior. Since different cycles do not intersect, to decide if cycle A is inside cycle B it is sufficient to check if a single point P of A lies inside B. This is accomplished by stereographically projecting the atom from the point P onto the plane tangent to the antipodal pont of P. This projection either preserves or reverses the orientation of cycle B according to whether P is inside or outside cycle B. Note that a cycle is oriented counter-clockwise when viewed from any point in its interior.

If a pair of cycles A, B are mutually interior, they belong to the same face provided any third cycle C which is interior to both A and B also has A and B in its interior. Using this final criterion the cycles can be partitioned according to the faces they border.

There are two cases where the situation can be simplified. First, if there is only one cycle on an atom, this bounds a single face which is a disc. Second, if a cycle consists of just one or two edges, then all other cycles lie in the interior of this one.

A.6. Shape sequences

The calculation of the shape sequence and the topological type of $\mathcal{S}$ are described in [19] and are included here for the sake of completeness. The procedure is to reassemble $\mathcal{S}$ from the visible faces of the atoms, one atom at a time, keeping track of the topology of the partially built surface along the way. At each stage, the partial surface may have several connected components, and the topology of each component is determined by the pair (χ, β), where χ is the Euler characteristic and β the number of boundaries of the component [12]. The sequence of pairs (χ, β), for each partial surface, is the *shape sequence* of the embedding [19].

Initially, the visible faces on the first atom form a collection of discs, each with $(\chi, \beta) = (1, 1)$. The faces on the other atoms are now added one by one, and the topological invariants updated as follows. Let $C(\Sigma)$ denote the set of components of a surface Σ. If a face F is joined to Σ, then $C(\Sigma \cup F)$ consists of $C(\Sigma)$ together with a new component $G = F \cup C_F$, where $C_F = \bigcup \{ c \in C(\Sigma) \colon c \cap F \neq \emptyset \}$. The edges of G are the edges of F plus the edges of each $c \in C_F$ minus those edges common to F and C_F. The number of boundaries of G is just the number of cycles amongst these edges, and the Euler characteristic of G is

$$\chi(G) = \chi(F) + \sum_{c \in C_F} \chi(c) - e + v,$$

where e and v are the number of edges and vertices common to F and C_F.

As the completed surface has no boundary, each pair in the final entry of the shape sequence has the form $(\chi, 0)$, and $\mathcal{S}$ has one component for each pair. If there is more than one such pair, and the solid model $\mathcal{M}$ is connected, this indicates the presence of one or more cavities within the solid model (cf. [1]). The topological type of each component of $\mathcal{S}$ is determined by its Euler characteristic χ or its genus $\gamma = (2 - \chi)/2$.

References

[1] P. Alard and S.J. Wodak, J. Comput. Chem. 12 (1991) 918.
[2] F. Aurenhammer, ACM Comput. Surv. 23 (1991) 345.

[3] D.R.J. Chillingworth, *Differential Topology with a View to Applications* (Pitman, London, 1976).

[4] M.L. Connolly, J. Appl. Crystallography 16 (1983) 548.

[5] G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation* (Research Studies Press, Taunton, 1988).

[6] H. Edelsbruner, *Algorithms in Combinatorial Geometry* (Springer, Berlin, 1987).

[7] E.L. Eliel, *Stereochemistry of Organic Compounds* (Wiley, Chichester, 1994).

[8] V. Guillemin and A. Pollack, *Differential Topology* (Prentice-Hall, NJ, 1974).

[9] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).

[10] C.A. James and D. Weininger, *Daylight Theory Manual* (Daylight Chemical Information Systems, Inc., 1995).

[11] D.G. Kendall, Bull. London Math. Soc. 16 (1984) 81.

[12] W.S. Massey, *Algebraic Topology: An Introduction* (Harcourt, Brace and World, New York, 1967).

[13] P.G. Mezey, *Shape in Chemistry: An Introduction to Molecular Shape and Topology* (VCH, New York, 1994).

[14] J. Milnor, *Morse Theory* (Princeton University Press, 1963).

[15] A. Okabe, B. Boots and K. Sugihara, *Spatial Tesselations: Concepts and Applications of Voronoi Diagrams* (Wiley, Chichester, 1992).

[16] D. Siersma, Preprint no. 967, Department of Mathematics, University of Utrecht (1996).

[17] N. Trinajstic, *Chemical Graph Theory*, Vols. I and II (CRC Press, Boca Raton, FL, 1983).

[18] D.C. Whitley, J. Chem. Inf. Comput. Sci., submitted.

[19] D.C. Whitley and M.G. Ford, in: *Molecular Similarity in Drug Design*, ed. P.M. Dean (Blackie, Glasgow, 1995) p. 269.

[20] R.J. Wilson, *Graphs: An Introductory Approach* (Wiley, Chichester, 1990).